



## King's Research Portal

DOI:

[10.1080/09638237.2017.1392008](https://doi.org/10.1080/09638237.2017.1392008)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Vitoratou, S., & Pickles, A. (2017). A note on contemporary psychometrics. *Journal of Mental Health*, 1-3. <https://doi.org/10.1080/09638237.2017.1392008>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# **A note on contemporary psychometrics**

Silia Vitoratou<sup>1,\*</sup> and Andrew Pickles<sup>1</sup>

<sup>1</sup>Psychometrics & Measurement Lab,  
Department of Biostatistics and Health Informatics,  
Institute of Psychiatry, Psychology, and Neurosciences,  
King's College London, UK

\*Corresponding author

Silia Vitoratou, PhD  
Lecturer in Psychometrics  
Psychometrics & Measurement Lab,  
Department of Biostatistics and Health Informatics  
IoPPN, King's College London,  
Room S2.10  
SE5 8AF, London, UK.  
Phone: 0207 848 0309  
Fax: 0207 848 0281  
Email: [silia.vitoratou@kcl.ac.uk](mailto:silia.vitoratou@kcl.ac.uk)

Psychometrics provide the mathematical underpinnings for psychological assessment. From the late 19th century, a plethora of methodological research achievements equipped researchers and clinicians with efficient tools whose practical value becomes more evident in the era of the internet and big data. Nowadays, powerful probabilistic models exist for most types of data and research questions. As the usability of the psychometric scales is better comprehended, there is an increased interest in applied research outcomes. Paradoxically, while the interest in applications for psychometric scales increases, publishing research on the development and/or evaluation of those scales per se, is not welcomed by many relevant journals. This special issue in psychometrics is therefore a great opportunity to briefly review the main ideas and methods used in psychometrics, and to discuss the challenges in contemporary applied psychometrics.

### **Main ideas in psychometrics**

In mental health sciences, as we cannot measure perceptions, emotions, attitudes, personality *traits* directly, we need to rely on their indirect assessment through the responses to a set of observed variables (items, questions, symptoms). These observed variables are often referred to as **indicators**, as they are designed to reflect the **latent trait** being assessed. But how do we learn about a trait via such observed items? Imagine looking out of a window towards a field of crops. Being indoors, one cannot feel or see the wind directly, yet by observing the movement of the crops, one can infer not only the existence of wind but also the wind's direction and magnitude. This is only possible because we naturally assume that the wind is the **common cause** which makes the crops move. Based on the movement of individual plants and as a group, we can measure the wind. Likewise, in psychometrics we assume that the latent trait affects the responses in our observed items and makes them vary. Therefore, based on our items' variation and covariation, we can measure their **common cause**, that is, the non-directly measurable trait.

Apart from the common cause, another idea essential for psychometrics, is the idea of **measurement error**. According to the classical test theory, our observed measurement is only partly the true value because our observation also includes random errors. The equation *observation equals true value plus error* is the core of classical test theory. This was the idea which prompted the field of psychometrics and provides the foundation for the definitions of **reliability** (precision) and **validity** (accuracy).

### **Psychometric properties and latent trait models**

Reliability refers to the question: ***are we measuring the trait with precision?*** It entails the reproducibility of the results, which is fundamental in all fields of science. Test-retest reliability (stability), inter-rater (equivalence) reliability, and internal consistency (equivalence/stability), all embody the classical test theory ideas for random error (each attributing it to different sources) and are estimated using standard statistical tests. The reliability of a scale is estimated specifically for a target population. A change in the target population from one setting to another (inpatient to community, for instance) will require further reliability assessments, evaluated independently.

With respect to validity, the question to be answered is: ***do we really measure what we intend to?*** The validity assessment is essentially our attempt to gather evidence to answer this question and we do so by utilising the broader scientific knowledge in our area of research. In this sense, the validity assessment of a scale is never actually finalised. It is always subject to new findings and understanding. Unlike reliability, validity does not address the random error; in fact, validity is employed to avoid systematic error. There are many different types of validity: face and content, predictive and concurrent, convergent, discriminant (or divergent), and discriminative, among others. Among the different types of validity, the importance of the face validity (*valid in the eyes of the target population*) is now widely comprehended (see also Neale and Strang, 2015). Patients' involvement in the item development stage has become a standard practice, improving the comprehension of the items and the psychometric properties of the instrument. This way the measurement reflects not only the clinician but also the patient experience. The validity assessment can be based in simple methods such as correlations and regressions, or extended to more sophisticated methods such as receiver operating characteristic curves to compare with gold-standard criteria or anchoring vignettes to address cross-cultural bias.

Knowledge of psychometric properties is vital in our attempt to minimise the measurement error. The measurement itself is achieved using multivariate statistical models, which belong to the family of the generalised latent trait models (Moustaki & Knott, 2000). The important task here is to choose the appropriate model for each application, based on the **type of data** and on the hypothesis under consideration. Binary data can be handled with the item response theory model or with item factor analysis, ordinal data with the partial credit or graded response models, continuous data with the common factor analysis model, exploratory and/or confirmatory. These models allow us to study the relationship of each item with the latent trait, to evaluate whether our trait is unidimensional or multidimensional, to address the measurement error and the psychometric properties in a sophisticated manner, and eventually to ***produce the measurement scores***.

More advanced models are used to study further important aspects of measurement, such as the **measurement bias**. This refers to the bias introduced to our measurement due to respondent characteristics such as gender, age, ethnicity, or conditions (different raters, re-assessments). To test for score differences between groups or conditions, we need first to make sure that the scale we use functions similarly in all cases. For instance, if our scale underestimates aggression in women, then any comparisons of their scores with scores from men is not to be trusted. Multiple group factor analysis models are used to ensure measurement invariance between groups. For continuous variables, such as age for instance, invariance can be evaluated using multiple responses-multiple indicators models (MIMIC). Longitudinal factor analysis can evaluate bias in time or in the presence of multiple raters.

Computing advances have facilitated the development of even more complex models, including the computationally demanding analogues under the Bayesian paradigm. Models can be fitted for either side of the debate: such as, “under what circumstances does it make sense to regard psychopathology as being scalar and under what circumstances does it make sense to regard psychopathology as being categorical” (Pickles and Angold, 2003).

### **Psychometrics in the literature**

There is an increased interest with respect to measurement. For instance, a recent literature review conducted by the Psychometrics and Measurement Lab (PML) at the IoPPN, revealed that *the number of publications in applied psychometrics affiliated with the institute almost doubled over the past five years*. At the same time, however, many funders will not support measure development work and many relevant journals appear to be reluctant to publish articles referring solely to newly developed scales. This discourages researchers from investing in the learning and application of modern psychometric methodology and in undertaking the systematic development work necessary to produce the best measurement tool for the task. This is disappointing, since the accomplishments of theoretical (methodological) research enable us to develop powerful psychometric instruments. We are now able to measure the traits of interest with precision and accuracy, to understand better their characteristics and properties, and to use this knowledge in clinical practice. Provided, however, that **the proper methods are used and that they are used properly**.

It is not rare in the literature, for instance, to see categorical data being analysed using common factor analysis (suitable for continuous data) leading to biased estimates and non-reproducible structures. Even when the correct method is used, following blindly the numerous indices and criteria, without a

proper understanding of their meaning, purpose, and interpretation, can lead to conflation or needless fragmentation of dimensions. Another problem is that whilst reliability, validity, and dimensionality are often studied and reported, measurement bias is commonly neglected. In a clinical setting, a consistent bias in measurement leads to over or under-diagnosis and consequent treatment in one group compared to another. A less than adequate evaluation of the properties of a scale can also lead to the artefactual generation of evidence for or against categorical or dimensional disorders, or even to inappropriate use of a measure designed for one setting or purpose, for another. These difficulties, along with the sustained flow of new methods provided by theoretical psychometrics and the widespread applications, call for **guidelines for researchers**, such as the COSMIN guidelines (Mokkink, 2010). Updated comprehensive methodological checklists are not only crucial for the development of new scales, but also vital in the *re-evaluation* and improvement of existing popular scales, in view of new methods and software. This will **improve the quality of the scales** eventually available to clinicians and researchers.

The Journal of Mental Health traditionally encourages and supports research in applied psychometrics (Cao et al, 2016; Fitzgerald et al, 2015; O'Connor et al, 2014; Roncalli et al, 2013; Lloyd et al, 2012, among others). This special issue adds to this effort and presents a variety of different research scenarios often occurring in applied psychometrics. Barber et al, (2017) and Newman-Taylor et al. (2017) present newly developed scales. Giromini et al, (2017) and Goossens et al, (2017) present the psychometric properties of pre-existing scales, which are currently translated in a different language. The process of translating a psychometric instrument is a rigorous one and entails all stages of evaluating a new instrument, apart from item development. Joshanloo & Jovanović, (2017) in addition to presenting a cross-cultural adaptation of a pre-existing scale, investigate the measurement invariance of the scale with respect to gender. Kocalevent et al, (2017) and Kim et al, (2017) on the other hand, examined the psychometric properties of pre-existing scales in different populations than the populations targeted when the scales were originally developed. Tomizawa et al, (2017) revise a pre-existing scale to adjust for multi-nationality, that is, to overcome cultural differences in the measurement. Wei et al, (2017) give an example of a systematic review of psychometric scales. Furnham & Crump (2017) give an example of using psychometric analysis to reduce uncertainty and finally Kelly presents a large cross-sectional psychometric study which spans in 21 European countries.

## Conclusion

Psychometrics is a valuable tool for psychological assessment in three distinct ways. First in everyday clinical practice, where standardised psychometric measures **add evidence** which can help the

clinician with diagnosis or formulation. Second in applied research, where psychometric scales offer a powerful tool for screening and/or evaluating the **prevalence of a certain trait in a population**, as a one-to-one expert assessment is not feasible. Third psychometrics provide the methods to measure, study, **understand, and explore a latent trait per se**, facilitating our effort to advance our knowledge in mental health sciences. Our raw data are complex, and though the methods too can appear overwhelming, they have the capacity to both justify and achieve great simplification. Developing a scale demands the co-evaluation of many results over an extended sequence of development stages, following both general and trait-specific, purpose-tailored, rules. There are no shortcuts, no generic or automated procedures to fit all applications and any available guidelines need to be adjusted to the task at hand. Exploring the magnificent world of identifying and measuring abstract concepts may be seeking simplicity but was never itself going to be simple.

## References

- Barber JM, Parsons H, Wilson CA, Cook CCH. (2017). Measuring mental health in the clinical setting: What is important to service users? The Mini-Service user Recovery Evaluation scale (Mini-SeRvE). J Ment Health, doi: 10.1080/09638237.2017.1340624 [Epub ahead of print]
- Cao J, Yang J, Zhou Y, et al. (2016). The effect of Interaction Anxiousness Scale and Brief Social Phobia Scale for screening social anxiety disorder in college students: a study on discriminative validity. J Ment Health, 25, 500–5
- Fitzgerald S, Umucu E, Arora S, et al. (2017). Psychometric validation of the Clubhouse climate questionnaire as an autonomy support measure for people with severe mental illness. J Ment Health, 24, 38–42
- Furnham A, Crump J. (2017). Personality correlates of passive-aggressiveness: A NEO-PI-R domain and facet analysis of the HDS Leisurely scale. J Ment Health, Available from: <http://dx.doi.org/10.3109/09638237.2016.1167853> [Epub ahead of print]

Giromini L, Colombarolli MS, Brusadelli E, Zennaro A. (2017). An Italian contribution to the study of the validity and reliability of the trait meta-mood scale. *J Ment Health*, doi: 10.1080/09638237.2017.1340621 [Epub ahead of print]

Goossens PJ, Beentjes TA, Knol S, et al. (2017). Investigating the reliability and validity of the Dutch versions of the illness management and recovery scales among clients with mental disorders. *J Ment Health*, Available from: <http://dx.doi.org/10.3109/09638237.2015.1124398> [Epub ahead of print]

Joshanloo M, Jovanović V. (2016). The factor structure of the Mental Health Continuum-Short Form (MHC-SF) in Serbia: An evaluation using exploratory structural equation modelling. *J Ment Health*, Available from: <http://dx.doi.org/10.1080/09638237.2016.1222058> [Epub ahead of print]

Kelly BD. (2017). Exploring and explaining the “Santa Claus effect”: Cross-sectional study of jollity in 21 European countries. *J Ment Health*, doi: 10.1080/09638237.2017.1370643 [Epub ahead of print]

Kim J. (2017). The factor structure of the Dispositional Hope Scale in hemiplegic stroke patients. *J Ment Health*, Available from: <http://dx.doi.org/10.1080/09638237.2017.1385735> [Epub ahead of print]

Kocalevent RD, Finck C, Pérez-Trujillo M, et al. (2017). Standardization of the Beck Hopelessness Scale in the general population. *J Ment Health*, Available from: <http://dx.doi.org/10.1080/09638237.2016.1244717> [Epub ahead of print]

Lloyd K, Devine P. (2012). Psychometric properties of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in Northern Ireland. *J Ment Health*, 21, 257–63

Mokkink LB, Terwee CB, Patrick DL, et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res*, 19, 539–49

Moustaki I, Knott M. (2000). Generalized Latent Trait Models. *Psychometrika*, 65, 391–411

Neale J, Strang J. (2015). Philosophical ruminations on measurement: Methodological orientations of patient reported outcome measures (PROMS). *J Ment Health*, 24, 123–5

Newman-Taylor K. (2017). Psychometric Evaluation of the Hope, Agency and Opportunity (HAO): A brief measure of mental health recovery. *J Ment Health*, doi: 10.1080/09638237.2017.1385746 [Epub ahead of print]



O'Connor M, Casey L, Clough B. (2014). Measuring mental health literacy – a review of scale-based measures. *J Ment Health*, 23, 197–204

Pickles A, Angold A. (2003). Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Dev Psychopathol*, 15, 529–51

Roncalli S, Byrne M, Onyett S. (2013). Psychometric properties of a Mental Health Team Development Audit Tool. *J Ment Health*, 22, 51–9

Tomizawa R, Yamano M, Osako M, et al. (2016). Validation of a global scale to assess the quality of interprofessional teamwork in mental health settings. *J Ment Health*, Available from: <http://dx.doi.org/10.1080/09638237.2016.1207232> [Epub ahead of print]

Wei Y, McGrath PJ, Hayden J, Kutcher S. (2017). Measurement properties of mental health literacy tools measuring help-seeking: A systematic review. *J Ment Health*, doi: 10.1080/09638237.2016.1276532. [Epub ahead of print]